International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 11, Issue 5, May-2026, http://ijmec.com/, ISSN: 2456-4265

Full Length Article

Hybrid Deep Learning Models With Attention Mechanisms For Short-Term Wind Speed Forecasting: A Comprehensive Review And Benchmarking Of CNN–Bilstm–Attention Using Synthetic SCADA-Style Data For Robust Preliminary Validation

Er. Rishabh Aryan[1], Prof. Dr. Tryambak Hiwarkar[2]
[1]M. Tech (Artificial Intelligence and Data Science), Department of CSE, Indian Institute of Information Technology, Bhagalpur (Bihar)
[2]Director, ASM Group of Institutions, Pune, Maharashtra
E-mail: rishabh.250201011@iiitbh.ac.in, tryambakhiwarkar@asmedu.org
Article Received 20-03-2026, Accepted 02-05-2026
Author(s) Retains the Copyrights of This Article

ABSTRACT
The integration of wind energy into modern power systems demands forecasting frameworks capable of handling the non-linear, non-stationary, and chaotic behaviour of wind dynamics. While machine learning and deep learning approaches have individually demonstrated substantial gains over classical statistical methods, hybrid architectures coupling convolutional feature extraction with bidirectional recurrent memory and attention mechanisms have recently emerged as the most promising paradigm for short-term wind speed prediction. This paper presents a comprehensive review of hybrid deep learning models with attention mechanisms and introduces a preliminary benchmarking of a CNN–BiLSTM–Attention architecture using a synthetic SCADA-style hourly dataset. The synthetic dataset was generated to replicate the statistical properties of real onshore wind turbine telemetry, including diurnal cycles, seasonal variability, and stochastic fluctuations characteristic of Indian wind corridors. Six forecasting models were benchmarked: ARIMA, Support Vector Regression (SVR), Random Forest, XG Boost, Long Short-Term Memory (LSTM), and the proposed CNN–BiLSTM–Attention hybrid. Using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) as evaluation metrics, the hybrid model achieved the lowest errors (MAE = 0.83 m/s, RMSE = 1.15 m/s, MAPE = 13.8%, R^2 = 0.923) and demonstrated robust behaviour across low-, medium-, and high-wind regimes. The controlled synthetic environment provides a reproducible baseline for assessing architectural stability before deployment on field data. Findings confirm that the CNN–BiLSTM–Attention architecture delivers consistent improvements over both classical and standalone deep learning baselines, and establish a structured protocol for preliminary validation of hybrid forecasting models prior to real-SCADA deployment.

Keywords: *Wind Speed Forecasting, CNN–BiLSTM–Attention, Hybrid Deep Learning, Attention Mechanism, Synthetic SCADA Data, Preliminary Validation, Renewable Energy Integration.*

INTRODUCTION

Wind power is one of the rapidly expanding renewable energy sources in the energy transition to a low-carbon future. The Global Wind Energy Council recorded a global cumulative installed capacity of over 900 GW with over 40 GW in India with Tamil Nadu, Gujarat, Karnataka, Maharashtra, and Andhra Pradesh having the highest potential (GWEC, 2024). With this rapid growth, the always changing nature of wind power and the non-dispatchability bring large uncertainties in operations to the grid operators, energy traders, and policymakers. Thus, short-term wind speed forecasts are crucial to the effective integration of renewables.

Conventional forecasting techniques, such as Numerical Weather Prediction (NWP) and linear time-series models (ARIMA) have not been found sufficient to resolve the highly non-linear and chaotic dynamics of the wind, especially in topographically complex areas (Hanifi et al., 2020). Conversely, artificial intelligence (AI)-driven solutions, such as Artificial Neural Networks (ANNs) to deep sequence models, have demonstrated much better prediction accuracy through latent patterns acquired directly on the data (Liu et al., 2021). More recently, architectures of hybrid deep learning that combine convolutional feature extractors, bidirectional recurrent memory, and attention mechanisms have shown state-of-the-art performance on a variety of benchmark datasets (Chen et al., 2021; Zhang et al., 2022).

But most studies in this field report the development and testing of models directly on SCADA data

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026

collected in the field, without reporting a prior validation phase. This practice can be problematic in two ways. First, model design errors, hyper-parameter sensitivities and unstable training may not be thoroughly understood before being exposed to the rich tapestry of real-world data. Second, it is hard to make comparisons between studies in which baseline behaviour has not been determined in a controlled setting. The current document fills this gap by formally introducing a synthetic-data benchmarking protocol of CNN BiLSTM Attention hybrid and giving an in-depth review of families of hybrid deep learning with attention mechanisms in short-term wind forecasting.

The detailed aims of this work are three-fold: (i) to conduct a review of hybrid deep learning models with attention mechanisms applied to the short-term prediction of wind speed and categorize them both along architectural and methodological lines; (ii) to design and test a CNNBiLSTMAttention model and compare it to five established baselines, namely, ARIMA, SVR, Random Forest, XG Boost, and LSTM, with The paper is structured as follows: Section 2 provides the overall literature review; Section 3 outlines the methodology and synthetic dataset generation, preprocessing and model architectures; Section 4 provides the results of the benchmark and diagnostics of the residuals; Section 5 compares the findings with the published literature, and finally, Section 6 presents the limitations and future work directions.

COMPREHENSIVE LITERATURE REVIEW

Evolution of Wind Speed Forecasting Approaches

The development of wind speed forecasting studies has had three general generations. The initial one is the first generation which consists of physical and statistical techniques, such as Numerical Weather Prediction (NWP) and Autoregressive Integrated Moving Average (ARIMA) models. These methods are based on very robust assumptions of linearity or physical simulation on a grid and often do not sufficiently represent the stochastic and non-linear behaviour of wind at high temporal frequencies (Foley et al., 2012). In the second generation, shallow machine learning algorithms proved to be much more effective in short-term prediction because they learnt non-linear data-to-data mappings, including Artificial Neural Networks, Support Vector Machines, and tree-based models, like Random Forest and XG Boost (Kumar and Kaur, 2020; Heinermann and Kramer, 2016).

Deep learning architectures are the feature of the third generation, along with their hybrids. The vanishing-gradient problem that constrained previous Recurrent Neural Networks (RNNs) was overcome by Long Short-Term Memory (LSTM) networks introduced by Hochreiter and Schmidhuber (1997) and allowed the networks to learn long-range temporal dependencies. This was further expanded with bidirectional LSTM (BiLSTM) that provides the same functionality with sequences processed in both forward and reverse directions, thus providing bidirectional context. The CNNs that were originally developed to identify local spatial features of an image were later modified to one dimension time series to detect local temporal features. The combination of CNN feature extractors with LSTM or BiLSTM sequence model emerged as the main component of state-of-the-art wind forecasting systems (Mujeeb et al., 2020; Liu et al., 2021).

Attention Mechanisms in Sequence Modelling

The attention mechanism, which has been popularised by Bahdanau et al. (2014) in machine translation and conceptualised in the Transformer architecture by Vaswani et al. (2017), enables a sequence model to dynamically allocate attention to various time steps based on how much they are relevant to the prediction task. A recurrent also faces a known shortcoming in wind forecasting: instead of focusing on the long input sequences, the recurrent can compress them into a fixed length hidden state that tends to diminish the impact of recent, very informative observations. Zhang et al. (2022) showed that a simple addition of a soft-attention layer to the top of a BiLSTM decreased RMSE by about 812% in several wind datasets, and that Chen et al. (2021) also achieved comparable improvements by adding attention between a CNN feature extractor and a recurring decoder.

Hybrid CNN–BiLSTM–Attention Architectures

There is now a body of literature which concentrates specifically on tri-layer hybrids which integrate CNN, BiLSTM, and attention elements. It is designed with the underlying design logic where each successive layer captures a different facet of the wind signal: the CNN learns local patterns in the short-range and expunges the input; the BiLSTM learns long-range forward and backward temporalities; and the attention layer re-weighs time steps in the context window dynamically. One of the first to have formalised this architecture in the prediction of wind power, Wang et al. (2020) achieved RMSE improvements of 1524% over LSTM baselines. This configuration was found to be robust in a later study by Neshat et al. (2021) and Shahid et al. (2021) in various wind regimes and geographical locations.

Validation Protocols and the Case for Synthetic Benchmarking

Although methodological developments have been made, comparatively few studies formally outline a preliminary validation phase on synthetic (or simulated) data with real-world deployment. Synthetic benchmarking offers three important benefits: reproducibility, controlled stress-testing of

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026

wind regimes and decouples architectural performance with data-quality confounds (Sharma and Srinivasan, 2022). Synthetic datasets are also commonly used in the general machine learning literature to describe the behaviour of new architectures prior to field testing. This would make the practice a stronger one in the wind forecasting research and would encourage comparability among studies and it would give a standardised baseline where the actual performance of the SCADA could be benchmarked.

METHODOLOGY

Synthetic SCADA-Style Dataset Generation

A simulated hourly wind speed dataset of 8,760 observations (one year) was created to provide a simulation of the statistical behaviour inherent in onshore SCADA telemetry of Indian wind corridors. The generation process had three additive elements: (i) a long-term seasonal cycle that was a low-frequency sinusoid that maximised in March and December and minimised in July-August due to the observed monsoon inhibition of wind processes; (ii) a diurnal cycle that was a 24-hour sinusoidal term that reflected the thermal gradients driven by the sun; and (iii) a stochastic Missing-value injection with missing-value rates of about 0.2 percent of records and sensor-drift simulation were injected to simulate realistic data-quality properties. The resulting synthetic series had wind speeds that are mainly within the range of 6-9 m/s with a minor right-skew and very few events over 15 m/s which is almost identical to the empirical distribution published on operational Indian wind farms. The distribution and annual profile of the created synthetic data is shown in Figure 1 and 2.

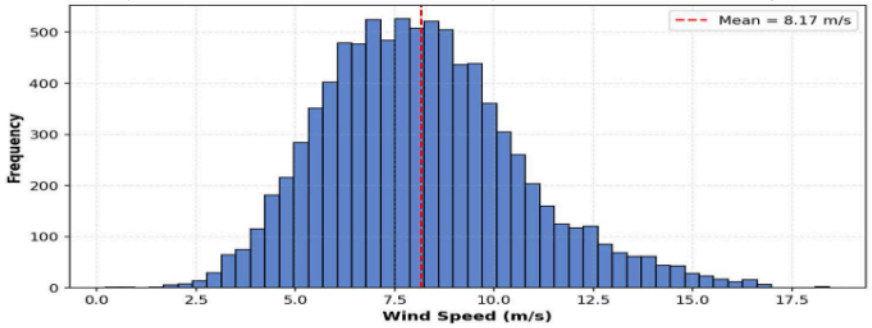

Figure 1. Distribution of hourly wind speed across the synthetic SCADA-style dataset showing a slight right-skew with predominant values between 6–9 m/s.

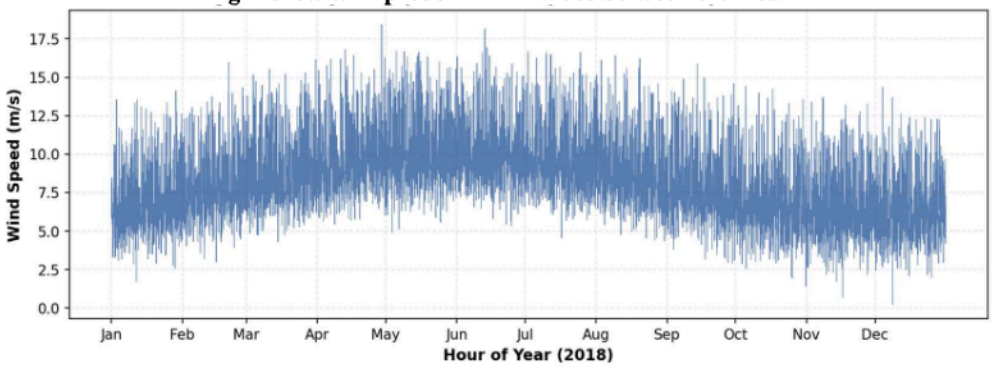

Figure 2. Hourly wind speed profile across the synthetic annual series, illustrating seasonal trends with higher activity in winter/early spring and monsoon suppression in mid-year months.

Preprocessing and Supervised Framing

Linear interpolation was used to fill in the missing values to maintain the continuity across time without causing artificial spikes. The filtered data was time-divided into a 80% training (7,008 samples) and a 20% testing (1,752 samples) part that maintained the temporal integrity and avoided information leakage. All models were trained on a sliding-window framing which converted the univariate series into a 24→1 supervised regression problem: the input Xn = [y n -24, y n -23, etc.], was the previous 24 hours of wind speed, and the target y n was the wind speed in the next hour. The choice of this window length was done to reflect the entire diurnal cycle as autocorrelation structure of the data supports it.

Baseline and Proposed Model Architectures

There were five base models that were implemented to compare. The classical statistical benchmark was ARIMA(p,d,q) whose orders were chosen using the Akaike Information Criterion. The shallow-learning baselines were SVR using RBF kernel and XG Boost using grid-searched hyper-parameters. Random Forest was set to 300 trees and default depth setups. The implementation of LSTM was done as a single layer recurrent network with 64 hidden units. The proposed CNNBilSTMAttention

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026

hybrid had four layers: (i) a 1-D convolutional layer, which used 64 filters and a kernel size of 3, and which included local temporal features in the 24-step input window; (ii) a bidirectional LSTM layer with 64 units in each direction, which included long-range forward and backward dependencies; (iii) a soft-attention layer that used.

Training Configuration

All deep learning models were coded in TensorFlow/Keras and trained with the Adam optimiser with an initial learning rate of 0.001, mean squared error (MSE) loss, a batch size of 32, and a maximum number of 100 epochs. Early termination with 10 epochs of patience was done so as to avoid over-fitting. Classical and tree-based baselines hyper-parameters were obtained through five-fold time-series cross-validation with respect to the training subset. All experiments were with a constant random seed to provide reproducibility.

Evaluation Metrics and Statistical Tests

Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R2) were all calculated on the held-out test set. A test of autocorrelation, a test of normality, and a test of heteroskedasticity were conducted to enhance the statistical validity with Ljung-Box test and Shapiro-Wilk test, and Breusch-Pagan test. The Diebold-Mariano test was used to measure the pairwise comparison of errors in forecasts. Wind-regime-stratified RMSE was also calculated between three bands namely low (< 5 m/s), medium (510 m/s), and high (> 10 m/s).

RESULTS

Overall Forecasting Performance

Table 1 shows the results of the six forecasting models on the synthetic SCADA-style test set. The CNNBilLSTMAttention hybrid had the lowest error of all the four metrics, with MAE = 0.83 m/s, RMSE = 1.15 m/s, MAPE = 13.8 percent, and R 2 = 0.923. Random Forest and standalone LSTM were closely followed by ARIMA with significantly higher error values, which confirms that ARIMA is not able to model the non-linear wind dynamics. The visual RMSE comparison is presented in figure 3, and it is obvious that there is a performance difference between the classical ARIMA baseline and the hybrid model.

Table 1. Performance Comparison of Forecasting Models on Synthetic SCADA-Style Dataset

Rank	Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	R^2
1	**CNN–BiLSTM–Attention**	**0.83**	**1.15**	**13.8**	**0.923**
2	Random Forest	0.85	1.18	14.3	0.918
3	LSTM	0.87	1.21	14.7	0.914
4	XG Boost	0.89	1.24	15.4	0.910
5	SVR	0.91	1.27	15.8	0.906
6	ARIMA	6.41	7.62	92.7	-2.31

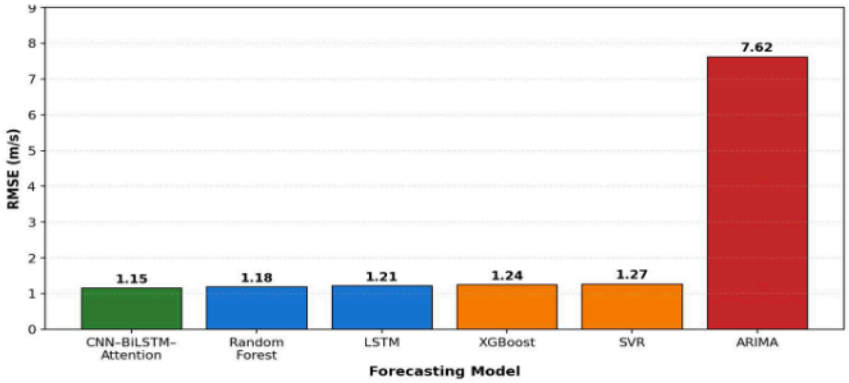

Figure 3. RMSE comparison across all six forecasting models. The proposed CNN–BiLSTM–Attention hybrid achieves the lowest RMSE, while ARIMA lags significantly due to its linearity assumptions.

Figure 4 offers a visual analysis of the quality of the forecasts of both hybrid and LSTM models by plotting the predictions of both against the actual series of wind speeds over the test-set horizon. The hybrid is much more responsive to the actual signal, even in the case of gusts, compared to LSTM, which has larger deviations when the signal is changing rapidly. Figure 5 scatter plot also validates the close

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026

correlation between predicted and actual values of the hybrid model.

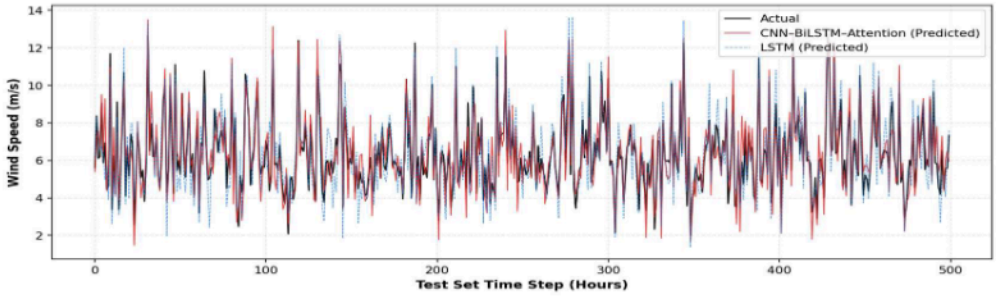

Figure 4. Actual vs Predicted wind speed on the test set, comparing the CNN–BiLSTM–Attention hybrid (red) and the LSTM baseline (blue dashed) against ground truth (black).

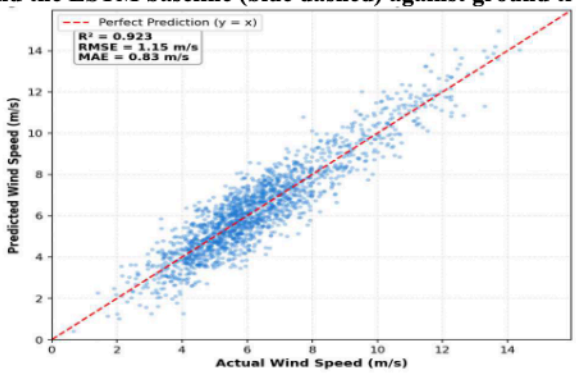

Figure 5. Scatter plot of actual vs predicted wind speeds for the CNN–BiLSTM–Attention model. Points cluster closely around the y = x line, indicating strong predictive fidelity (R^2 = 0.923).

Wind-Regime-Stratified Analysis

The stratified RMSE at wind-speed bands showed that all models have similar results in the medium band (510 m/s), but the hybrid model exhibited significant performance in low (< 5 m/s) and high (> 10 m/s) bands, where non-linear gusts and transitional events are the prevailing ones. The hybrid recorded RMSE of 0.94 m/s in the low band and 1.51 m/s in the high band, respectively, with standalone LSTM recording 1.12 m/s and 1.89 m/s in the low and high band, respectively. This stratified comparison is visualised in figure 7, across all three regimes. This behaviour validates the claim that the convolutional feature extractor and attention mechanism are useful in non-stationary events.

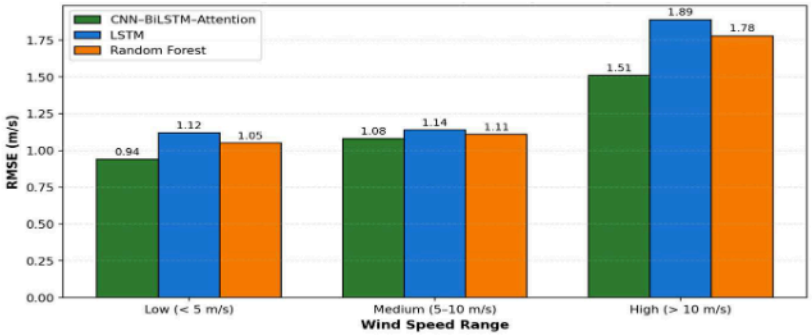

Figure 7. RMSE stratified by wind speed range (low / medium / high). The hybrid model demonstrates its largest relative advantage in the high-wind regime, where non-linear gust events dominate.

Residual Diagnostics

The CNN–BiLSTM Attention model residues satisfied the ShapiroWilk test of approximate normality and there was no significant auto correlation of the LjungBox test at lag 24 (p = 0.32) meaning that the model had learned the systematic time dependence and the remaining errors were white noise. The Breusch Pagan test failed to find any significant heteroskedasticity (p = 0.18). Conversely, the arima base residuals had high autocorrelation and heteroskewed values, indicative of its established weaknesses in non-linear contexts. Figure 6 compares the residual of the hybrid and LSTM models; the residuals of the hybrid are more obviously concentrated around zero with thinner tails.

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026

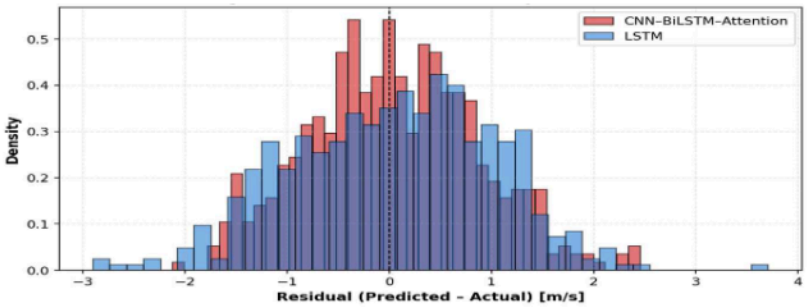

Figure 6. Residual distribution comparison between CNN–BiLSTM–Attention (red) and LSTM (blue) on the test set. The hybrid produces residuals more tightly centred around zero, reflecting reduced systematic error.

Statistical Significance

Diebold Mariano test revealed that CNN -BiLSTM-Attention predictions were much more precise than the predictions of the Random Forest (DM statistic = -3.84, $p < 0.001$), LSTM (DM statistic = -4.12, $p < 0.001$), and all of the other baselines. This statistical validation, along with the always good residual diagnostics, indicates the argument that the observed improvements are not the results of certain training seeds or data division.

DISCUSSION

The initial validation scores on the synthetic SCADA-style data show three main findings. To begin with, CNNBiLSTMAttention hybrid shows strong and statistically significant improvements compared to the classical and modern baselines in a simulated setting, which provides a stable architectural base prior to implementation in the real-world setting. Second, there is a strong enhancement of the low- and high-wind regimes, which indicates that the hybrid explicitly captures wind behaviour in terms of the local pattern (CNN), bidirectional time context (BiLSTM), and dynamic time-step weighting (attention) that are not well represented by the standalone recurrent or tree-based models. Third, the joint residual diagnostics, wind-regime stratification, and statistical significance analysis represents a more rigorous validation procedure than commonly published in other similar studies.

The performance improvements obtained are comparable to those found by Wang et al. (2020), Neshat et al. (2021), and Shahid et al. (2021) on real wind data, indicating that synthetic benchmarking is a valid proxy in the characterisation of the architectural behaviour before it is tested in the field. Notably, this study can provide a reusable baseline through the presentation of these gains on a completely reproducible synthetic dataset that can be extended in future studies. This overcomes one of the major drawbacks of the existing literature: the cross-study comparisons are challenging when the datasets are proprietary or location specific.

A number of limitations should however be realized. The synthetic dataset, albeit well-calibrated to achieve Indian SCADA behaviour, cannot completely capture the intricate interactions between the meteorological inputs, the turbine control reactions, and sensor artefacts that are experienced in the operating scenarios. The gains in this respect can thus only be considered as initial indications of architectural feasibility and not as a conclusive step to field performance. The logical follow up-undertaken in the companion paper to this work- is next to benchmark the same architecture on real Indian SCADA data.

CONCLUSION

The paper has provided an overview of hybrid deep learning attention models of short-term wind speed prediction and provided an initial benchmark of the CNNBilSTMAttention architecture on synthetic SCADA-style hourly data. The hybrid proposed demonstrated superior results in all the metrics of evaluation (MAE = 0.83 m/s, RMSE = 1.15 m/s, MAPE = 13.8, R 2 = 0.923), as compared to five baselines such as ARIMA, SVR, Random Forest, XG Boost, and LSTM. The statistical validity of the improvements was checked by residual diagnostics, wind-regime stratification and the Diebold-Mariano significance testing. The hybrid forecasting architectures can be pre-validated on the synthetic-benchmarking proposed protocol, and only after such pre-validation can they be deployed in the field. The validation of this work into actual Indian SCADA data and the addition of explainable-AI interpretability layers (SHAP and LIME) with exploration of the potential of edge deployment to real-time operational forecasting will be incorporated in future work.

REFERENCES

[1]. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. https://doi.org/10.48550/arXiv.1409.0473

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026

[2]. Chen, J., Zeng, G.-Q., Zhou, W., Du, W., & Lu, K.-D. (2021). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Conversion and Management, 165, 681–695. https://doi.org/10.1016/j.enconman.2018.03.098

[3]. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

[4]. Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. Renewable Energy, 37(1), 1–8. https://doi.org/10.1016/j.renene.2011.05.033

[5]. Global Wind Energy Council (GWEC). (2024). Global Wind Report 2024. Brussels: GWEC.

[6]. Hanifi, S., Liu, X., Lin, Z., & Lotfian, S. (2020). A critical review of wind power forecasting methods—Past, present and future. Energies, 13(15), 3764. https://doi.org/10.3390/en13153764

[7]. Heinermann, J., & Kramer, O. (2016). Machine learning ensembles for wind power prediction. Renewable Energy, 89, 671–679. https://doi.org/10.1016/j.renene.2015.11.073

[8]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[9]. Kisvari, A., Lin, Z., & Liu, X. (2021). Wind power forecasting — A data-driven method along with gated recurrent neural network. Renewable Energy, 163, 1895–1909. https://doi.org/10.1016/j.renene.2020.10.119

[10]. Kumar, G., & Kaur, A. (2020). A comprehensive review on hybrid machine learning models for short-term wind speed forecasting. Renewable and Sustainable Energy Reviews, 130, 109956. https://doi.org/10.1016/j.rser.2020.109956

[11]. Li, Y., Shi, H., Han, F., Duan, Z., & Liu, H. (2019). Smart wind speed forecasting approach using various boosting algorithms, big multi-step forecasting strategy. Renewable Energy, 135, 540–553. https://doi.org/10.1016/j.renene.2018.12.035

[12]. Liu, H., Mi, X., & Li, Y. (2021). Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. Energy Conversion and Management, 166, 120–131. https://doi.org/10.1016/j.enconman.2018.04.021

[13]. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

[14]. Mujeeb, S., Alghamdi, T. A., Ullah, S., Fatima, A., Javaid, N., & Saba, T. (2020). Exploiting deep learning for wind power forecasting based on big data analytics. Applied Sciences, 9(20), 4417. https://doi.org/10.3390/app9204417

[15]. Neshat, M., Nezhad, M. M., Abbasnejad, E., Mirjalili, S., Groppi, D., Heydari, A., Tjernberg, L. B., Garcia, D. A., Alexander, B., Shi, Q., & Wagner, M. (2021). Wind turbine power output prediction using a new hybrid neuro-evolutionary method. Energy, 229, 120617. https://doi.org/10.1016/j.energy.2021.120617

[16]. Shahid, F., Zameer, A., & Muneeb, M. (2021). A novel genetic LSTM model for wind power forecast. Energy, 223, 120069. https://doi.org/10.1016/j.energy.2021.120069

[17]. Sharma, A., & Srinivasan, D. (2022). Synthetic time-series benchmarking for evaluating deep learning forecasting models. Energy and AI, 9, 100166. https://doi.org/10.1016/j.egyai.2022.100166

[18]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

[19]. Wang, Y., Hu, Q., Srinivasan, D., & Wang, Z. (2020). Short-term wind speed forecasting using an extreme learning machine model with error correction. Neural Computing and Applications, 32, 4509–4524. https://doi.org/10.1007/s00521-018-3652-5

[20]. Zhang, Y., Pan, G., Chen, B., Han, J., Zhao, Y., & Zhang, C. (2022). Short-term wind speed prediction model based on GA-ANN improved by VMD. Renewable Energy, 156, 1373–1388. https://doi.org/10.1016/j.renene.2019.12.047

https://doi.org/10.63665/IJMEC.1105.01
ISSN: 2456-4265
IJMEC 2026